Griko, Y. V., Privalov, P. L., Venyaminov, S. Y., & Kutyshenko, V. P. (1988) *J. Mol. Biol. 202*, 127–138.
Hanson, J. C., & Schoenborn, B. P. (1981) *J. Mol. Biol. 153*, 117–146.
Harrison, S. C., & Blout, E. R. (1965) *J. Biol. Chem. 240*, 299–303.
Hughson, F. M., & Baldwin, R. L. (1989) *Biochemistry 28*, 4415–4422.
Kawamura-Konishi, Y., Kihara, H., & Suzuki, H. (1988) *Eur. J. Biochem. 170*, 589–595.
Kumar, A., Ernst, R. R., & Wüthrich, K. (1980) *Biochem. Biophys. Res. Commun. 95*, 1–6.
La Mar, G. N., Pande, U., Hauksson, J. B., Pandey, R. K., & Smith, K. M. (1989) *J. Am. Chem. Soc. 111*, 485–491.
Lecomte, J. T. J., & Cocco, M. J. (1990) *Biochemistry* (preceding paper in this issue).
Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967–974.
Privalov, P. L., Griko, Y. V., Venyaminov, S. Y., & Kutyshenko, V. P. (1986) *J. Mol. Biol. 190*, 487–498.
Radding, J. A. (1987) *Biochemistry 26*, 3530–3536.
Rance, M., & Wright, P. E. (1986) *J. Magn. Reson. 66*, 372–378.
Rance, M., Sørensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., & Würthrich, K. (1983) *Biochem. Biophys. Res. Commun. 117*, 479–485.
Takano, T. (1977a) *J. Mol. Biol. 110*, 537–568.
Takano, T. (1977b) *J. Mol. Biol. 110*, 569–584.
Teale, F. W. J. (1955) *Biochim. Biophys. Acta 35*, 289.
Waltho, J. P., Feher, V. A., & Wright, P. E. (1990) in *Current Research in Protein Chemistry* (Villafranca, J. J., Ed.) Section III, pp 283–293, Academic Press, New York.

Refolding and Aggregation of Bovine Carbonic Anhydrase B: Quasi-Elastic Light Scattering Analysis†

Jeffrey L. Cleland and Daniel I. C. Wang*
Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139
Received April 16, 1990; Revised Manuscript Received August 28, 1990

ABSTRACT: Bovine carbonic anhydrase B (CAB) is chosen as the model protein to study the phenomenon of protein aggregation, which often occurs during the refolding process. Refolding of CAB from 5 M GuHCl has been observed by quasi-elastic light scattering (QLS), which confirms the formation of a molten globular protein structure as reported previously [Semisotnov, G. V., Rodionova, N. A., Kutyshenko, V. P., Ebert, B., Blanck, J., & Ptitsyn, O. B. (1987) *FEBS Lett. 224*, 9–13]. QLS analysis reveals the formation of multimeric species prior to precipitation. Activity and cross-linking studies have confirmed the presence of inactive multimeric protein species. The dimer formation has been determined to be the initiating step in the aggregation of CAB during refolding. Activity studies have indicated that the first intermediate observed in the refolding pathway of CAB aggregates to form the inactive dimer. The rate of formation of the dimer has a stoichiometric dependence on the final protein concentration. The dimer formation rate is a function of the final guanidine hydrochloride (GuHCl) concentration to the inverse 6.7 power, which correlates well with the binding of GuHCl to the native protein in 0.60–0.80 M GuHCl. These rate dependencies require the refolding of CAB to be performed at high GuHCl concentrations (1 M GuHCl) and low protein concentrations (less than 1 mg/mL) to avoid the formation of aggregates. Alternatively, refolding can be performed by allowing the first intermediate to form the second intermediate prior to further dilution or dialysis. The aggregation of a hydrophobic first intermediate species is likely to be common to the refolding of other molten globular proteins.

Purification of recombinant proteins expressed in *Escherichia coli* often requires the solubilization and renaturation of proteins that are expressed as insoluble inclusion bodies (Marston, 1986). During the refolding of proteins from denatured states, aggregation of partially refolded protein occurs resulting in decreased recovery of the native protein. Early studies of protein refolding by Anfinsen and others have indicated that many proteins aggregate during refolding from unfolded states (Anfinsen & Haber, 1961; Epstein & Goldberger, 1963). The aggregation of protein during refolding is strongly dependent on the final protein and denaturant concentration (Zettlmeissl et al., 1979). Therefore, proteins are usually refolded at very low protein concentrations (μg/mL) or high denaturant concentrations (1 M guanidine hydrochloride or 4 M urea) to avoid aggregation.

To investigate the phenomenon of protein aggregation during refolding, the study of the model protein system, bovine carbonic anhydrase B (CAB),[1] is undertaken since CAB has a well-characterized refolding pathway (Stein & Henkens, 1978; Doligkh et al., 1984; Semisotnov et al., 1987). During refolding of CAB from its denatured state in 5 M GuHCl, the protein will rapidly form a compact molten globule structure with exposed hydrophobic clusters. These clusters will then

† This work was supported by National Science Foundation Grant CDR-88-03014.
* To whom correspondence should be addressed.

[1] Abbreviations: CAB, bovine carbonic anhydrase B; GuHCl, guanidine hydrochloride; QLS, quasi-elastic light scattering; Tris, tris(hydroxymethyl)aminomethane; EDTA, ethylenediaminetetraacetic acid; pNPA, *p*-nitrophenol acetate; pNP, *p*-nitrophenol; DMS, dimethyl suberimidate dihydrochloride; SDS, sodium dodecyl sulfate; PAGE, polyacrylamide gel electrophoresis; CONTIN, constrained regularization method; SDP, size distribution processor.

0006-2960/90/0429-11072$02.50/0

collapse to form a core in the second intermediate. Finally, the native protein structure of CAB is formed with an overall half-time of 12 min (Stein & Henkens, 1978; Semisotnov et al., 1987). CAB also aggregates during refolding at high protein concentrations (mg/mL) and low denaturant concentrations (0.1–0.7 M GuHCl) (Ikai et al., 1978).

To study the kinetics of aggregate formation during refolding, QLS is used to measure conformational changes as well as the appearance of submicron protein aggregates. Previous studies have applied classical light scattering to measure the kinetics of protein aggregate formation during refolding (Zettlmeissl et al., 1979). However, QLS can effectively analyze the early stages of protein aggregation where submicron aggregation predominates. The measurement of initial protein crystallization and protein–protein interactions has also been studied by utilizing QLS techniques (Donovan et al., 1987; Yarmush et al., 1988; Mikol et al., 1989; Kadima et al., 1990). The inherent advantages of this technique include noninvasive analysis and rapid measurement, which facilitates the study of aggregation kinetics. The analysis of CAB refolding and aggregation utilizing QLS will provide insight into the kinetics of protein aggregation and assist in the development of models for aggregation leading to improvements in protein refolding.

Experimental Procedures

Materials

Bovine carbonic anhydrase B (CAB), bovine serum albumin, guanidine hydrochloride (GuHCl), Tris–sulfate, ethylenediaminetetraacetic acid (EDTA), *p*-nitrophenol acetate (pNPA), and ammonium acetate were molecular biology grade and purchased from Sigma Chemical Co. (St. Louis, MO). The purity of the CAB (pI = 5.9) was checked by gel electrophoresis and silver staining. Dimethyl suberimidate dihydrochloride (DMS) was obtained from Pierce (Rockford, IL). Triethanolamine was purchased from Mallinckrodt (Paris, KY). HPLC grade acetonitrile and acetone were from J.T. Baker (Phillipsburg, NJ). The SDS-PAGE materials were obtained from Pharmacia LKB Biotechnology (Uppsala, Sweden). All buffers and samples were prepared with distilled water passed through a MilliQ water purification system (Millipore Corp., Bedford, MA).

Methods

Protein Concentration. Protein concentration for native CAB in 50 mM Tris–sulfate and 5 mM EDTA, at pH 7.5 was determined by absorbance at 280 nm with an extinction coefficient of 1.83 (mg/mL protein)$^{-1}$ cm^{-1} (Wong & Tanford, 1973) and a molecular weight of 30000. For CAB denatured in 5 M GuHCl, the concentration was determined by using a colorimetric dye binding assay (Bio-Rad Laboratories, Richmond, CA) using bovine serum albumin denatured in 5 M GuHCl as the standard.

Quasi-Elastic Light Scattering (QLS) Measurements. All QLS measurements were performed using a Model N4 submicron particle analyzer (Coulter Electronics, Hialeah, FL) instrumented as described previously (Yarmush et al., 1988). The photomultipler assembly was positioned at 90° to the incident laser beam (5-W argon ion, Model 2020-05, Spectra Physics, Mountain View, CA). Samples were analyzed at a constant temperature of 20 °C with a total volume of 200 μL. All buffers and protein solutions were prefiltered with 0.22-μm syringe filters (Gelman Sciences, Ann Arbor, MI) to remove any dust particles that would alter the QLS measurements. Glassware was rinsed in acetone and filtered distilled water to remove excess dust. QLS sample tubes (6 × 50 mm, VWR Scientific, San Francisco, CA) were checked for imperfections prior to use.

The methods of determining particle size distributions from QLS autocorrelation function data have been well documented (Stock & Ray, 1985). The method of constrained regularization or CONTIN as described by Stock and Ray was used in the analysis of all autocorrelation data. The Model N4 system deploys a size distribution processor (SDP) system that utilizes CONTIN to calculate particle size distributions (Coulter Electronics, 1984). The SDP and CONTIN analysis methods were utilized for CAB refolding and aggregation measurements.

For each renaturation experiment, the CAB was first denatured in 5 M GuHCl for at least 6 h. The denatured protein was then measured by QLS to ascertain that only the unfolded random coil existed in solution. A mean hydrodynamic diameter of 10 nm was obtained for each denaturation. The theoretical hydrodynamic diameter of the unfolded protein was calculated as 10.4 nm using the reported intrinsic viscosity (Wong & Tanford, 1973; Corbett & Roche, 1984). Refolding was carried out by rapid dilution with a dilution buffer containing 50 mM Tris–sulfate and 5 mM EDTA, pH 7.5, to the desired final protein and GuHCl concentrations. The sample was analyzed by QLS immediately after dilution. The earliest possible time point obtained by QLS measurements was 30 s, which is the minimum time needed for the autocorrelation channels to stabilize. The sample time used to collect autocorrelation data for each experiment was 30 s and the midpoint of this time period was used to determine the rate of association. For more rapid association kinetics, a sample time of 20 s was used to obtain the autocorrelation data. Each experiment was repeated a minimum of three times with the same time points to assure repeatability and to reduce the difficulties of signal to noise. Continuous measurements were made during the refolding until an equilibrium state was achieved or until micron-sized aggregates dominated in the sample. Each experiment was repeated several times to provide a significant number of time points for kinetic analysis.

Multimer Model of Submicron Aggregates. To determine the unit composition of the particles initially formed during aggregation, a model of the multimeric state of the protein was developed on the basis of the wormlike chain model (Kratky & Porod, 1949). This model has been shown to successfully characterize the composition of antigen–antibody complexes that were analyzed by QLS (Murphy, 1989; Murphy et al., 1990). For purposes of this study, the multimers were assumed to consist of individual unit monomers. The unit monomer hydrodynamic diameter, D_h, was determined from the hydrodynamic diameter of the native structure at the same final protein and denaturant concentrations. The native protein structure was observed to change under different final concentrations and over long time periods to eventually reach the same equilibrium state as the refolding experiments (see Results). Therefore, if the measured native protein hydrodynamic diameter was greater than that of the first observed intermediate, the diameter of the first intermediate was employed in the calculation of multimer size. For each experiment, the mean hydrodynamic radius of the monomer, native or first intermediate, was used as the persistence length in the wormlike chain model.

Two multimeric states were calculated utilizing the chain model. First, the dimer was calculated assuming a contour length, L_i, ranging from two contacting spheres with 25% of their volume overlapped ($1.5D_h$) and a linear chain of solid spheres with no excluded volume ($2D_h$). The trimer of CAB

was the other multimer species that was assumed to range in contour length from a three-sphere cluster ($2D_h$) to a linear chain of solid spheres ($3D_h$). In addition, the linear form of the trimer was considered energetically more favorable than the triangular aggregates as determined from previous studies on the potential energy relations of native protein aggregates (Reithel, 1963). Higher order multimers were not evaluated and modeled since the maximum observed diameter did not exceed that of the linear trimer species. Since the true multimer diameters are not known, a normal distribution about the mean diameter was used to calculate the actual concentration of the multimeric species (Murphy, 1989; Murphy et al., 1990). The validity of this model for CAB association has been recently confirmed by studies using size-exclusion chromatogaphy (Cleland & Wang, 1990). These chromatography studies yield the same multimer concentrations as those calculated by using the wormlike chain model. The particle structure factor calculated from other models also does not vary significantly over the range of hydrodynamic radii observed in these experiments (Murphy, 1989). Therefore, alternative model systems may yield similar multimer concentrations.

Esterase Activity. An enzymatic activity assay was performed by using the esterase reaction as described previously (Pocker & Stone, 1967). The unfolded CAB in 5 M GuHCl was rapidly diluted to the desired final protein and GuHCl concentrations and each aliquot of the refolded sample was analyzed for its enzymatic activity at various times during aggregation. Each assay sample was diluted 10-fold by 50 mM Tris–sulfate and 5 mM EDTA, pH 7.5, prior to addition of substrate, pNPA. The formation of pNP and decrease in pNPA were measured by absorbance at 348 and 400 nm, respectively, on a Model 8452 diode array spectrophotometer (Hewlett-Packard, Mountain View, CA) at 1-s time intervals for 2 min after addition of pNPA. Recovery of activity was determined using the ester hydrolysis rate constant of the native protein at the same concentration in the dilute buffer (50 mM Tris–sulfate, 5 mM EDTA, pH 7.5).

Cross-Linking Analysis. To confirm the validity of the multimer model, cross-linking experiments were performed during the refolding process. The cross-linking agent, DMS, was found to have no effect on the native protein under the same final protein and denaturant conditions. Denatured CAB in 5 M GuHCl was rapidly diluted with buffer (50 mM Tris–sulfate, 5 mM EDTA, pH 7.5) to a final protein and denaturant concentration as desired for the experiments. DMS at a final concentration of 20 mM was added after a given time period to cross-link the multimers (Swaney & O'Brien, 1978; Hadju et al., 1976). After 1 min, the cross-linking was quenched by the addition of 1 M ammonium acetate to yield a final concentration of 100 mM. Further quenching was also performed by the addition of concentrated GuHCl to yield a final concentration of 5 M GuHCl, which prevented further aggregation. Samples were analyzed by SDS–PAGE using 8–25% gradient polyacrylamide gels with SDS buffer strips on a Phast gel electrophoresis sytem (Pharmacia LK Biotechnology, Uppsala, Sweden). The gels were stained with Coomassie Blue by using the development section of the Phast System.

Results

For each refolding experiment, a compact initial structure was immediately formed upon dilution of the unfolded protein in 5 M GuHCl to low GuHCl concentrations (less than 1.0 M) as shown in Figure 1. The mean hydrodynamic diameter of this structure was approximately the same as that of the native protein molecule (D_h = 4.8 ± 0.5 nm). This initial molten globule intermediate was the aggregating unit monomer from which the multimer formation was derived. As shown in Figure 1, the protein grew in hydrodynamic diameter over time until micron-sized aggregates dominated the scattering in the solution. A similar effect was observed in a study of the native protein at various protein and GuHCl concentrations and a significant change in the size distribution was observed for conditions where multimers were observed to form. The addition of concentrated GuHCl to the native protein resulted in the formation of a structure that was larger than the native protein in the absence of GuHCl (Figure 1). The formation of this structure was caused by the destabilization of the tertiary structure of the protein as reported previously (Rodionova et al., 1989). After equilibration in GuHCl concentrations of less than 1 M, the native protein was found to have a particle size larger than that of the unfolded protein in 5 M GuHCl (D_h = 9.8 ± 0.8 nm). In many cases, the native protein formed micron-sized aggregates after extended periods of exposure to moderate GuHCl concentrations (0.30–0.80 M). These results were used to confirm the formation of submicron aggregates or multimers and provide information on the unit

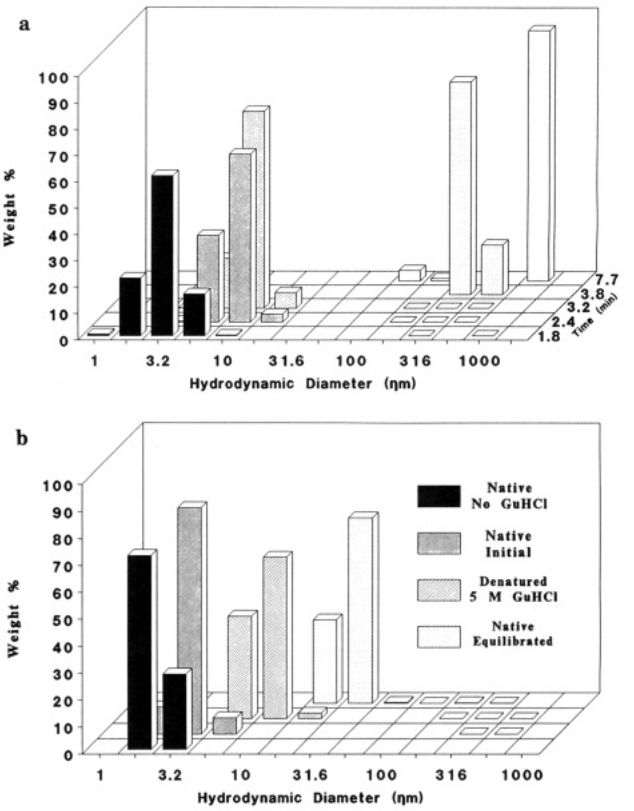

FIGURE 1: Size distributions for refolding, native, and denatured states of CAB. The *z*-average weight percent of each particle is shown over a range of hydrodynamic diameter as determined by CONTIN analysis of the QLS results. (A) Size distribution of CAB during refolding from 5 M GuHCl is shown as a function of time. CAB is refolded by rapid dilution to a final protein concentration of 0.50 mg/mL and a final GuHCl concentration of 0.70 M. The size distribution increases with time to form micron-sized aggregates after 3 min. (B) Native CAB size distribution dependence on GuHCl concentration is shown as the control for all refolding experiments. The native protein is mixed with concentrated GuHCl to yield a final protein concentration of 0.50 mg/mL and 0.60 M GuHCl. This mixture is immediately measured by QLS and the distribution is labeled as native initial. After equilibration for several hours under these conditions, the native protein in GuHCl is measured by QLS. The distribution after equilibrium is labeled native equilibrated. The size distributions of the native protein without GuHCl and the denatured protein in 5 M GuHCl are also included for reference.

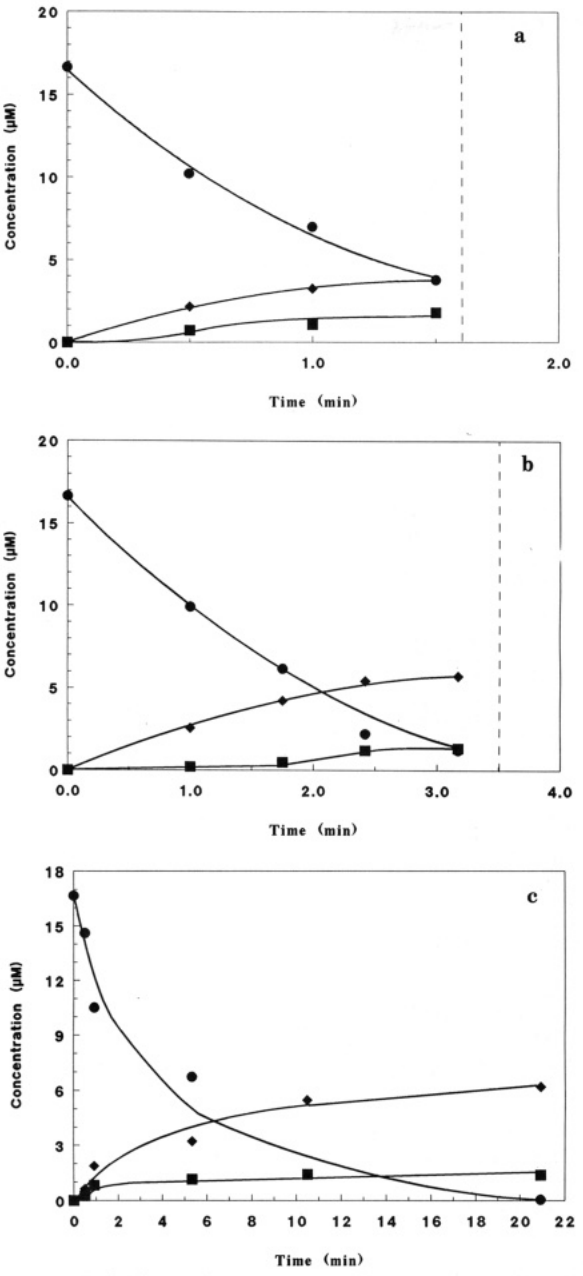

FIGURE 2: Refolding and aggregation of CAB as observed by QLS. The changes in monomer (●), dimer (◆), and trimer (■) concentration are displayed as a function of time after dilution for each case. (A) Rapid formation of multimers is measured where CAB in 5 M GuHCl is rapidly diluted to a final protein concentration of 0.50 mg/mL (16.7 μM) and 0.60 M GuHCl. Micron-sized aggregates (- - -) appeared after 1.5 min. (B) Moderate rate of multimer formation occurs for rapid dilution of CAB in 5 M GuHCl to 0.50 mg/mL (16.7 μM) and 0.70 M GuCHl. After 3.5 min, micron-sized aggregates (- - -) dominate the light scattering in the solution. (C) Multimers form slowly for rapid dilution of CAB in 5 M GuHCl to the final conditions of 0.50 mg/mL (16.7 μM) and 0.80 M GuHCl. Under these conditions, no further aggregation was observed after 20 min, indicating an apparent equilibrium state.

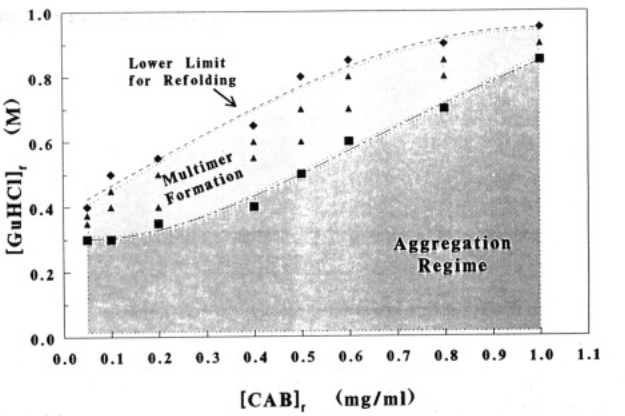

FIGURE 3: Regimes of refolding and aggregation of CAB. Each data point represents rapid dilution of CAB in 5 M GuHCl to a given final protein and GuHCl concentration. The aggregation regime is defined as the final solution conditions that result in the immediate formation of micron-sized aggregates. The upper boundary of the aggregation regime is defined by the lower data points (■). As depicted in Figure 2, the cases where dimer and trimer species are observed prior to micron-sized aggregation constitute the multimer formation regime (▲). The lower limit of refolding is the distinct regime where multimers form but do not proceed to form micron-sized aggregates (◆). From the lower limit of refolding to 1 M GuHCl, the protein refolds to form either a stable intermediate or the native structure in the absence of aggregation.

monomer hydrodynamic diameter.

After determination of unit monomer size distribution, multimer concentrations for each CAB refolding study were calculated on the basis of the QLS results with the exception of cases where aggregation to micron-sized particles occurred immediately upon dilution of the unfolded protein in 5 M GuHCl. For conditions of low final GuHCl concentrations (0.20–0.70 M GuHCl), CAB was observed to form precipitates during the dilution step. The next distinct type of aggregation was the rapid formation of multimers prior to micron-sized aggregate formation (Figure 2). When unfolded CAB in 5 M GuHCl was rapidly diluted to the final conditions of 0.50 mg/mL (16.7 μM) and 0.60 M GuHCl, the protein concentration decreased to 0.11 mg/mL (3.8 μM) at 1.5 min with the subsequent formation of the dimer and trimer species. For this case, micron-sized aggregates dominated the scattering after 1.5 min. The onset of micron-sized aggregates was delayed by increasing the final GuHCl concentration to 0.70 M at the same final protein concentration (0.50 mg/mL, 16.7 μM). Rapid dilution of denatured CAB in 5 M GuHCl to 0.50 mg/mL protein and 0.70 M GuHCl showed a decreased rate of multimer formation (Figure 2). In this case, the final observed monomer concentration before large aggregate formation was significantly reduced (0.04 mg/mL, 1.2 μM). The monomer concentration was observed to proceed to zero in the absence of large aggregates when the GuHCl was further increased to 0.80 M at the same final protein concentration (0.50 mg/mL, 16.7 μM). After dilution of the denatured CAB in 5 M GuHCl to 0.50 mg/mL (16.7 μM) and 0.80 M GuHCl, multimers formed relatively slowly and achieved a stable state or equilibrium after approximately 15 min (Figure 2). This stable equilibrium persisted for several hours and did not result in micron-sized aggregates. Several other final protein and denaturant conditions yielded the same type of distinct stages of multimer formation and micron-sized aggregation.

By combining the results of several different experiments, an operating diagram was constructed indicating the distinct regimes that occurred at different final CAB and GuHCl concentrations (Figure 3). In each case, the denatured protein in 5 M GuHCl was rapidly diluted to the final conditions shown in Figure 3. The aggregation regime was considered to be the region with final conditions that resulted in aggregation of CAB to micron-sized aggregates immediately upon dilution and, therefore, could not be analyzed by QLS. At higher final GuHCl concentrations, the multimer species as shown in Figure 2 were observed before the formation of large aggregates. The final condition where dimer and trimer achieved an equilibrium in the absence of larger aggregates (Figure 2) was considered the lower limit of refolding where

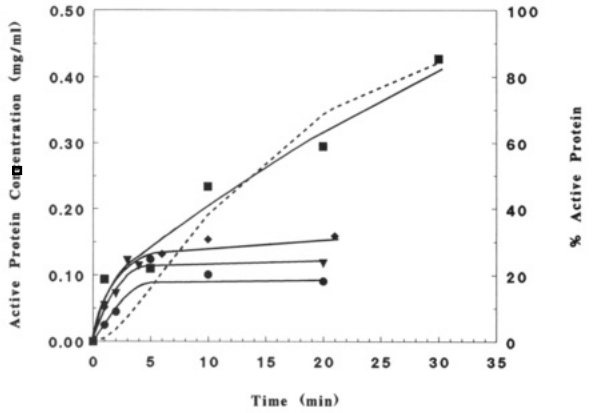

FIGURE 4: Concentration of active CAB as a function of time after rapid dilution from 5 M GuHCl. In each case, the protein is diluted to a final concentration of 0.50 mg/mL. The final GuHCl concentration for each case is 0.60 M (●), 0.70 M (▼), 0.80 M (◆), and 1.0 M (■). For the conditions where micron-sized aggregates are formed (0.60 and 0.70 M GuHCl), the accuracy of the assay is greatly reduced due to light scattering. The dashed line represents the concentration of native protein predicted on the basis of the rate constants from reported refolding experiments with an overall half-time of 12 min [refolding from 3 M GuHCl to 1 M GuHCl by rapid dilution with 0.05 M Tris buffer and 2 M NaCl, pH 7.5 at 20 °C (Stein & Henkens, 1978); refolding from 8.5 M urea to 4.1 M urea by rapid dilution with 0.1 M Tris buffer, pH 8 at 20 °C (Semisotnov et al., 1987)].

higher GuHCl concentrations at the same protein concentration resulted in refolding. The maximum GuHCl concentration to allow refolding was observed to be 1 M, which is comparable to results of CAB refolding reported by others (Dolgikh et al., 1984).

The recovery of activity of CAB was used to confirm the results of the QLS studies and determine the extent of refolding. Reactivation of denatured CAB in 5 M GuHCl was performed by rapid dilution to a given final protein and GuHCl concentration. At a final protein concentration of 0.50 mg/mL, the recovery of active CAB was complete after 5 min for final GuHCl concentrations from 0.60 to 0.80 M (Figure 4). For refolding at 0.50 mg/mL protein and 0.80 M GuHCl, maximum renaturation was 30% active protein, indicating the formation of inactive protein species. To confirm that the protein would refold at higher GuHCl concentrations, refolding was performed at 1.0 M GuHCl and 0.50 mg/mL protein. At 1.0 M GuHCl, recovery of activity was greater than 80% after 30 min. These results correlate well with reported CAB refolding kinetics as indicated by the dashed line in Figure 4 (Stein & Henkens, 1978; Semisotnov et al., 1987). The measured active protein concentration at the initial time points is greater than the predicted value since the second intermediate in the refolding pathway contains approximately 25% of the native enzyme activity (Wong & Tanford, 1973; Rodionova et al., 1989). The final activity for GuHCl concentrations of 0.60 and 0.70 M was greater than expected due to the presence of large aggregates, which affect the absorbance measurements for samples taken at times where micron-sized aggregates were observed by QLS. For a final GuHCl of 0.80 M, the dimers and trimers partially dissociated when diluted prior to the addition of substrate as observed by QLS and transmission electron microscopy (results not shown). This dilution effect explains the lack of correlation between the final activity and the final monomer concentration for the low GuHCl experiments (0.60–0.80 M GuHCl).

To determine if submicron multimers were forming prior to precipitation, cross-linking studies were performed to essentially lock the protein in the multimeric state. For refolding

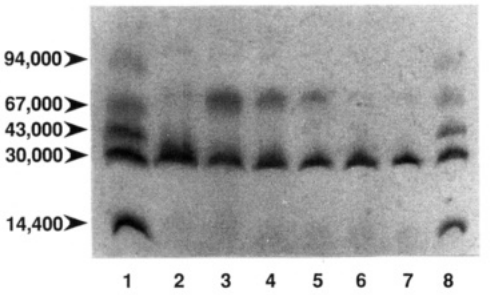

FIGURE 5: SDS–PAGE results of cross-linking during CAB refolding and aggregation. DMS is added to a solution of CAB that has been diluted from 5 to 0.70 M GuHCl with a final protein concentration of 0.50 mg/mL (see the Experimental Procedures section for details). Each sample is brought to a GuHCl concentration of 5 M by addition of 8 M GuHCl. Lanes 1 and 8 are low molecular weight markers in 5 M GuHCl (MW 14 400, 30 000, 43 000, 67 000, and 94 000). Lanes 2 and 7 contain the starting denatured protein solution in 5 M GuHCl. Lane 3 represents the results of cross-linking a refolding solution (0.50 mg/mL CAB, 0.70 M GuHCl) at 1.50 min. The results of cross-linking the refolding solution at 1.0 and 0.5 min are shown in lanes 4 and 5, respectively. Lane 6 is the control where the native protein is subject to 0.70 M GuHCl and the cross-linking agent at a protein concentration of 0.50 mg/mL.

of CAB from 5 M GuHCl to 0.70 M GuHCl and 0.50 mg/mL protein, the formation of dimers was observed to occur and the dimer band intensity increased with the refolding time (Figure 5). The increase in dimer concentration with refolding time was similarly predicted by the QLS analysis as shown in Figure 2. The trimer species was probably not observed in this experiment due to the low sensitivity of the staining technique, which can detect typically 0.1 mg/mL protein (Pharmacia LKB Biotechnology, Uppsala, Sweden). When the cross-linked protein solution was not increased to 5 M GuHCl immediately after quenching, large aggregates were rapidly formed, indicating the stabilization of an aggregating species.

Since the refolding pathway of CAB has been well characterized, it is possible to determine which intermediate protein structure forms the dimer and larger aggregates. Unfolded CAB in 5 M GuHCl rapidly folds into a compact molten globule structure when diluted to 1 M GuHCl (Doligkh et al., 1984). The unfolded protein does not exist in solution after 200 ms and the first intermediate with exposed hydrophobic clusters is completely converted to the second intermediate after 15 min (Semisotnov et al., 1987). The propensity of exposed hydrophobic groups on proteins to cause aggregation suggests that the first intermediate is the structure that forms the dimer. To confirm this hypothesis, refolding of denatured CAB in 5 M GuHCl has been performed by dilution to conditions where no aggregation occurs at 0.83 mg/mL protein and 1 M GuHCl and was followed by incubation for 15 min to assure that the first intermediate is no longer present. After 15 min, this solution is diluted to conditions (0.50 mg/mL, 0.60 M GuHCl) that usually result in aggregation as shown in Figure 2. The final solution, 0.50 mg/mL protein and 0.60 M GuHCl, did not form dimers or aggregates as observed by QLS (mean hydrodynamic diameter of 4.0 nm). In addition, this process of CAB refolding results in the same recovery of activity as protein refolded to 0.50 mg/mL and 1 M GuHCl (Figure 6). These results suggest that the first intermediate is the monomer responsible for the formation of dimers and larger aggregates during the refolding of CAB.

Discussion

The results of both QLS and cross-linking analyses indicate that a dimer species exists before the formation of micron

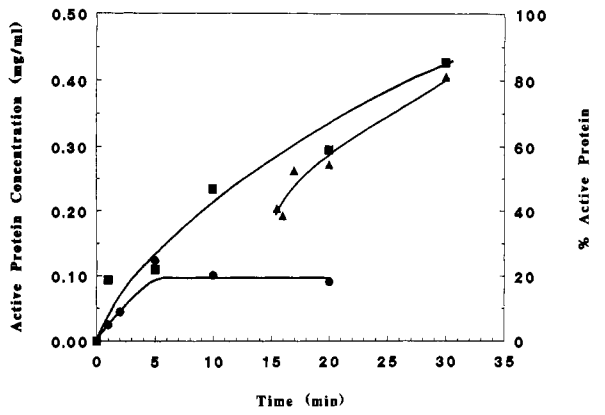

FIGURE 6: Aggregating species determination. The concentration of active protein over time during the refolding of CAB is shown for three different cases. The protein in 5 M GuHCl is rapidly diluted to 0.50 mg/mL and 0.60 M GuHCl (●) or 1.0 M GuCHl (■). To determine if the aggregating species is the first intermediate, the protein is first diluted to 0.83 mg/mL and 1.0 M GuHCl. After 15 min, the solution is diluted further to 0.50 mg/mL and 0.60 M GuHCl (▲) (see the Results section for details).

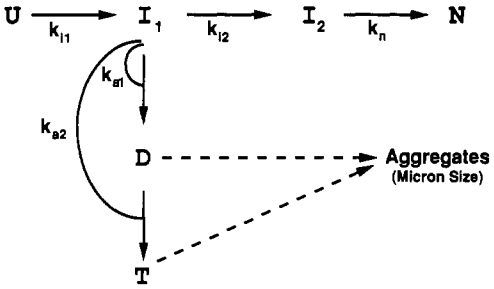

FIGURE 7: Proposed pathway for refolding and aggregation of CAB. The unfolded protein (U) rapidly forms the first compact intermediate (I_1) within 200 ms (k_{i1} = 23.1 s^{-1}). The first intermediate can then proceed to form the second intermediate (I_2, k_{i2} = 0.297 min^{-1}) or the dimer species (D) where k_{a1} is a function of the final protein and GuHCl concentration. The second intermediate continues to refold to form the native protein structure (N, $k_n = 6.93 \times 10^{-2}$ min^{-1}) (Stein & Henkens, 1978; Semisotnov et al., 1987). The dimer may form the trimer species (T) with the addition of another intermediate protein (I_1) to the aggregate. The rate constant, k_{a2}, is also a function of the final protein and GuHCl concentrations. The dimer, trimer, and first intermediate will then form the micron-sized aggregates under aggregating conditions (see Figure 3).

aggregates (Figures 2 and 5). In addition, the renaturation studies indicate that CAB does not completely recover activity in the absence of precipitation (Figures 2 and 4). The lack of activity of the protein in the absence of precipitation can be explained by a stabilization of inactive or partially active protein structures. These structures are the dimer and, possibly, trimer of inactive CAB, both of which are dissociated by further dilution in the activity analysis. These multimeric species do not form in the absence of the first intermediate as shown in Figure 6 and, therefore, the first intermediate is the monomer that associates to form the multimers and larger aggregates.

With the knowledge of the aggregating species and the formation of submicron aggregates, a refolding and aggregation model for CAB is postulated as shown in Figure 7. The model is based on rapid dilution of CAB from 5 M GuHCl to refold the protein. The formation of dimers, trimers, and large aggregates will occur when the protein is diluted to the final conditions of multimer formation or aggregation as depicted in Figure 3. The unfolded protein (U) rapidly forms the first intermediate (I_1) with a rate constant, k_{i1}, of 23.1 s^{-1}. The first intermediate will then either proceed on the refolding pathway to form the second intermediate (I_2) with a rate constant, k_{i2}, of 0.297 min^{-1} or form the dimer species with a rate constant, k_{a1}, that is dependent upon the final protein and GuHCl concentrations. Finally, the second intermediate will slowly fold into the native conformation (N) with a rate constant, k_n, of 6.93×10^{-2} min^{-1} (Semisotnov et al., 1987). The slow formation of the native structure is probably caused by the cis–trans isomerization of the 19 proline residues in CAB. Proline isomerization has been discussed as the possible cause of slow refolding in other protein systems (Kim & Baldwin, 1982). The formation of trimers and large aggregates is also included in the model, but their rates of formation and associated pathways have not been studied in this research.

Table I: Initial Rate Analysis of Dimer Formation for CAB Aggregation[a]

$[CAB]_f$ (mg/mL)	$[CAB]_f$ (μM)	$[GuHCl]_f$ (M)	R_D (μM/min)	$[D]_f$ (μM)	t_f (min)
0.50	16.7	0.60	5.22	3.76	1.5
0.50	16.7	0.70	3.68	5.72	3.2
0.50	16.7	0.80	0.74	6.22	20.9
0.60	20.0	0.80	0.84	4.93	9.5
0.80	26.7	0.80	2.63	4.40	5.2

[a] Determination of dimer concentration described under Experimental Procedures.

The kinetics of formation of the dimer species can now be studied on the basis of the proposed model of aggregation and refolding for CAB. The initial rate of dimer formation (R_D) for several different final conditions is shown in Table I. With a constant final GuHCl concentration, the rate of dimer formation is determined to have the following relationship to protein concentration:

$$R_D = k_d[CAB]^{2.6} \quad (1)$$

which approximately represents two molecules in the first intermediate state associating to form the dimer. Since there is an obvious dependence on the final GuHCl concentration at a constant final protein concentration, the rate constant, k_d, must be a function of the GuHCl concentration. Therefore, the rate of dimer formation as function of the final GuHCl concentration has been determined for a constant protein concentration (0.50 mg/mL). The dimer formation rate depends on GuHCl to the inverse 6.7 power, resulting in the following overall equation:

$$R_D = k'_d[CAB]^{2.6}[GuHCl]^{-6.7} \quad (2)$$

This strong inverse dependence on GuHCl may be explained by the predicted binding of GuHCl to the protein. For 0.50 mg/mL CAB in 0.60–0.80 M GuHCl, the number of GuHCl molecules bound to the protein should range from seven to nine from relationships for GuHCl binding to native proteins (Lee & Timasheff, 1974; Arakawa & Timasheff, 1984). Finally, the initiation of precipitating aggregates may be a nucleation and growth phenomena that has been observed previously for protein crystallization (Reithel, 1963; Feher & Kam, 1985). The final observed dimer concentration $[D]_f$, and the time of the final QLS measurement, t_f, are illustrated in Table I. For each case, the final observed dimer concentration is consistently between 4 to 6 μM, which is probably the critical nuclei concentration required for further aggregation. Additional studies of the formation of precipitating aggregates must be performed to determine if this aggregation can be modeled as a nucleation and growth process.

Conclusion

Quasi-elastic lighting scattering (QLS) has been demonstrated as a useful tool in the study of protein refolding and

aggregation. QLS analysis combined with renaturation and cross-linking studies has provided insight in the possible pathway for aggregation of CAB during refolding. The model developed in this work indicates that the refolding of CAB proceeds through an intermediate species, the first observed intermediate, which will aggregate under the appropriate conditions (Figures 3 and 7). To avoid aggregation, one must therefore refold CAB at high GuHCl concentrations or low protein concentrations. Alternatively, the refolding process can be performed by dilution to conditions that do not cause aggregation and incubation for 15 min, followed by additional dilution to the desired final conditions as demonstrated in Figure 6.

The model of CAB refolding and aggregation may also be common to several other proteins. Several proteins have been observed to form a molten globular intermediate structure during refolding (Vondervisz et al., 1987; Brems & Havel, 1989; Kuwajima, 1989; Garvey et al., 1989). In particular, dihydrofolate reductase and bovine growth hormone have been observed to form a compact molten globule intermediate (Garvey et al., 1989; Brems & Havel, 1989). For these two proteins, this intermediate structure contains exposed hydrophobic clusters analogous to the intermediate observed to form dimers and larger aggregates in the refolding of CAB. The model of aggregation and refolding for these proteins would likely be similar to that postulated for CAB, but the final protein and GuHCl concentrations required to obtain aggregation may be different on the basis of the difference in protein structure and binding of GuHCl. The generality of this model will be studied by QLS analysis of other protein systems.

With an understanding of the pathway for aggregation, research is currently under way to develop general methods to prevent aggregation by the inhibition of the dimer formation and to verify the proposed model for CAB refolding and aggregation. To confirm that the dimer and trimer species exist prior to precipitation, transmission electron microscopy (TEM) and additional cross-linking studies will be performed. Further characterization of the mechanisms of this aggregation process will be accomplished by spin labeling the protein for electron spin resonance (ESR) analysis, which has been done previously for CAB refolding (Semisotnov et al., 1987). These results should provide additional insight into the mechanism of protein aggregation during refolding and aid in the development of methods to avoid aggregation.

ACKNOWLEDGMENTS

The authors thank Professor Clark Colton, Department of Chemical Engineering, MIT, and Professor Martin Yarmush, Department of Chemical and Biochemical Engineering, Rutgers University, for the use of their Coulter N4 light scattering equipment as well as many helpful discussions. In addition, we especially thank Professor Peter Kim, Department of Biology, Whitehead Institute, MIT, for his useful suggestions and guidance throughout this work.

Registry No. CAB, 9001-03-0.

REFERENCES

Anfinsen, C. B., & Haber, E. (1961) *J. Biol. Chem. 236*, 1362–1363.

Arakawa, T., & Timasheff, S. N. (1984) *Biochemistry 23*, 5924–5929.

Brems, D. N., & Havel, H. A. (1989) *Proteins: Struct., Funct., Genet. 5*, 93–95.

Cleland, J. L., & Wang, D. I. C. (1990) in *Protein Refolding* (Georgiou, G., Ed.) ACS Symposium Series, American Chemical Society, Washington, DC (in press).

Corbett, R., & Roche, R. (1984) *Biochemistry 23*, 1888.

Doligkh, D. A., Kolomiets, A. P., Bolotina, I. A., & Ptitsyn, O. B. (1984) *FEBS Lett. 165*, 88–92.

Donovan, J. M., Benedek, G. B., & Carey, M. C. (1987) *Biochemistry 26*, 8116–8125.

Epstein, C. J., & Goldberger, R. F. (1963) *J. Biol. Chem. 238*, 1380–1383.

Feher, G., & Kam, Z. (1985) *Methods Enzymol. 114*, 49–76.

Garvey, E. P., Swank, J., & Matthews, C. R. (1989) *Proteins: Struct., Funct., Genet. 6*, 259–266.

Hadju, J., Bartha, F., & Friedrich, P. (1976) *Eur. J. Biochem. 68*, 373–383.

Ikai, A., Tanaka, S., & Noda, H. (1978) *Arch. Biochem. Biophys. 190*, 39–45.

Kadima, W., McPherson, A., Dunn, M. F., & Jurnak, F. A. (1990) *Biophys. J. 57*, 125–132.

Kim, P. S., & Baldwin, R. L. (1982) *Annu. Rev. Biochem. 51*, 459–489.

Kratky, O., & Porod, G. (1949) *Recl. Trav. Chim. 68*, 1106–1122.

Kuwajima, K. (1989) *Proteins: Struct., Funct., Genet. 6*, 87–103.

Lee, J. C., & Timasheff, S. N. (1974) *Biochemistry 13*, 257–265.

Marston, F. A. O. (1986) *Biochem. J. 240*, 1–12.

Mikol, V., Hirsch, E., & Giegé, R. (1989) *FEBS Lett. 258*, 63–66.

Murphy, R. M. (1989) Ph.D. Thesis, Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139.

Murphy, R. M., Yarmush, M. L., & Colton, C. K. (1990) *Biopolymers* (submitted for publication).

Pocker, Y., & Stone, J. T. (1967) *Biochemistry 6*, 668–678.

Reithel, F. J. (1963) *Adv. Protein Chem. 18*, 124–226.

Rodionova, N. A., Semisotnov, G. V., Kutyshenko, V. P., Uverskii, V. N., Bolotina, I. A., Bychkova, V. E., & Ptitsyn, O. B. (1989) *Mol. Biol. (Moskow) 23*, 683–692.

Semisotnov, G. V., Rodionova, N. A., Kutyshenko, V. P., Ebert, B., Blanck, J., & Ptitsyn, O. B. (1987) *FEBS Lett. 224*, 9–13.

Stein, P. J., & Henkens, R. W. (1978) *J. Biol. Chem. 253*, 8016–8018.

Stock, R. S., & Ray, W. H. (1985) *J. Polym. Sci., Polym. Phys. Ed. 23*, 1393–1447.

Swaney, J. B., & O'Brien, K. (1978) *J. Biol. Chem. 253*, 7069–7077.

Vondervisz, F., Lakatos, S., Gâl, P., Sârvâri, M., & Zâvodszky, P. (1987) *Biochem. Biophys. Res. Commun. 148*, 92–98.

Wong, K.-P., & Tanford, C. (1973) *J. Biol. Chem. 248*, 8518–8523.

Yarmush, D., Murphy, R. M., Colton, C. K., Fisch, M., & Yarmush, M. L. (1988) *Mol. Immunol. 25*, 17–32.

Zettlmeissl, G., Rudolph, R., & Jaenicke, R. (1979) *Biochemistry 18*, 5567–5571.